(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 August 2001 (02.08.2001)

PCT

(10) International Publication Number
WO 01/55911 A1

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number: PCT/US01/02527

(22) International Filing Date: 26 January 2001 (26.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/491,920 27 January 2000 (27.01.2000) US

(71) Applicant: INFORMAX, INC. [US/US]; 6110 Executive Boulevard, 10th Floor, North Bethesda, MD 20852 (US).

(72) Inventor: BABENKO, Vadim; Informax, Inc., 6110 Executive Boulevard, 10th Floor, North Bethesda, MD (US).

(74) Agents: ROBERTS, Jon, L. et al.; Roberts Abokhair and Mardula, LLC, Suite 1000, 11800 Sunrise Valley Drive, Reston, VA 20191 (US).

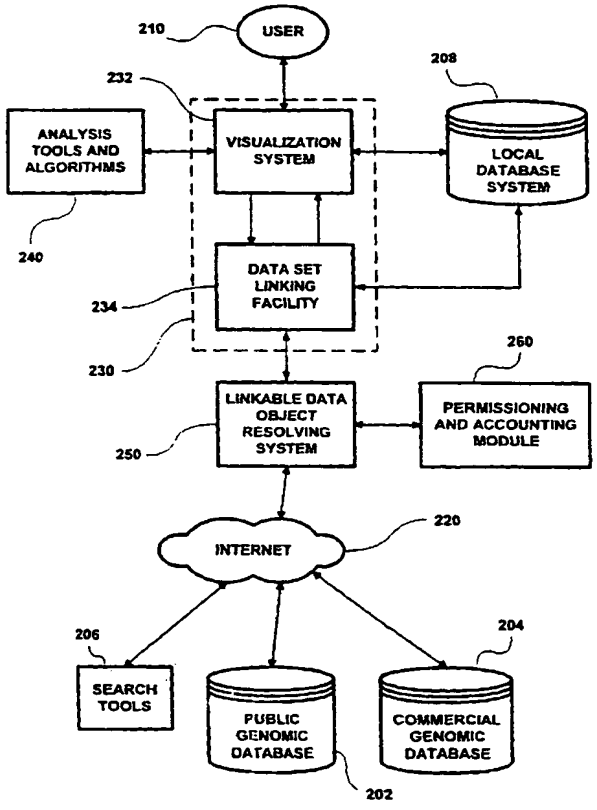
(81) Designated States *(national)*: DE, GB, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: INTEGRATED ACCESS TO BIOMEDICAL RESOURCES



(57) Abstract: A system and method of performing genomic research via online analysis of genomic data objects. A novel technical information model and implementation facility combines genomic visualization and analysis tools with the ability to link genomic data objects across (and within) plural databases. This provides for interactivity that is not possible using conventional Internet access processes and also provides for data integration that further enhances the effectiveness of genomic research. A business process controls how the databases are accessed, particularly the commercial fee-for-access databases, so that the researcher using a system according to the invention need not be concerned with keeping track of access limitations.

WO 01/55911 A1

Title: **INTEGRATED ACCESS TO BIOMEDICAL RESOURCES**

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to the field of data processing. More particularly, the present invention relates to the use of the Internet to access a variety of biomedical data resources in an integrated fashion, in order to provide for increased efficiency in genomic research.

2. Background Information

Almost every cell of every living organism contains a complete set of instructions for creating that organism and regulating its cellular structures and activities over its lifetime. That set of instructions is called a genome.

A genome is organized into distinct, microscopic units called chromosomes. Chromosomes are coiled threads of deoxyribonucleic acid (DNA). Each thread of DNA is composed of two long chains of nucleotides bound together in pairs to form a double helix. The human genome is made up of three and a half billion of these nucleotide pairs.

A given DNA strand contains the cells' instructions for producing proteins. These instructions are in the form of specific sequences of nucleotide bases, called genes, within the DNA strand. Scientists estimate that 80,000 to 100,000 of these basic units of heredity exist within the human genome. Proteins perform a wide variety of physiological tasks. They facilitate processes such as digestion, breathing, immune responses, the production of heat and energy, and the movement of fluids in and out of cells.

Most members of a species have the same collection of genes. However, each individual's unique characteristics stem from slight variations in the sequence of the

nucleotides that comprise the genes of that individual. These slight genetic variations that define unique characteristics of individuals are called polymorphisms. On average, the DNA of any two individuals in a species will differ by about 0.1%.

Another class of variations -- called mutations -- also occurs. Both polymorphic and mutagenic variations may be harmful to an individual by inhibiting the production, or altering the normal function, of a protein. Most diseases result from these types of genetic variations.

Genomics is the study of the nucleic sequences within a genome. The goal of genomic inquiry is to identify the sequence of nucleotides, understand the function of every gene they comprise, and clarify the genetic variations that define individuality and create disease.

Genomics has a broad scope of applications. These range from the most basic of research endeavors to the promise of diagnostic usefulness.

An important factor limiting the development of new drugs is the limited number of known target molecules for which new drugs can be developed. Disease target molecules are those that can be affected by a drug and cause a subsequent, desired biological reaction in the body. Historically, the process of discovering new target molecules has been extremely slow and very expensive due to reliance on trial-and-error approaches to discovery. Genomic research will reduce the reliance on trial-and-error by enabling drug designers to go directly to target molecules of interest. Thus, applying genomic research to drug development should produce new and better drugs more quickly, and at a reduced cost.

Another way that genomic research can help the pharmaceutical industry is in the emerging field of pharmacogenomics. Pharmacogenomics focuses on identifying genetic

variation among patients that may affect the efficacy of drug treatment – how well an individual's body absorbs and metabolizes a specific drug -- in order to develop more personalized drug therapies. Nearly all drug companies are developing pharmacogenomics units as a reaction to increasing evidence that a given drug does not have the same effect on all people.

In particular, pharmacogenomics is believed to offer at least three different useful applications:

- Increasing the success rate of clinical trials by improving the process of patient population selection;
- Identifying new uses for existing drugs; and
- Rescuing drugs that have failed previous drug trials by identifying more appropriate populations for using the drug. Candidates for drugs to be rescued include those that produce adverse reactions in particular sub-populations.

Molecular toxicology is another area of technology that can benefit from genomics research. Approximately 2.2 million Americans are admitted to hospitals every year as a result of adverse side effects from drugs. Over 100,000 Americans die annually from these adverse (and often unpredictable) effects. For instance, some cause liver damage, while others are harmful to the kidneys. Organ-specific gene expression profiles for drugs already available will enable researchers to study the toxicity of new drug compounds with more certainty.

In addition, gene expression data, combined with polymorphism information related to metabolic pathways, will provide important indications of the way an individual patient will react to drugs of various dosage levels, thereby significantly reducing the unwanted side effects of therapy.

Risk assessment is a major area of diagnostics that will benefit from genomics. Historically, prediction of whether someone is at special risk for a particular disease has focused on measuring general indicators in the body, such as blood pressure and cholesterol levels. These measurements reflect general physiology but do not explain the specific genetic basis of disease in an individual patient. Consequently, these diagnostic tests do not discern the underlying cause of disease and can result in compromised medical care for patients and increased risk of litigation.

New genomic-based diagnostics will focus on determining an individual's risk of developing a particular disease by looking at specific genes and any disease-related changes in that patient. These new diagnostics will likely lead to far better preventive care by offering more accurate assessments of a patient's potential risk for developing a particular disease.

Personalized medicine is another major area of diagnostics that will benefit from genomics. Genomic information will be available to develop molecular diagnostic tests to identify the genetic make-up of individuals. These diagnostic tests will revolutionize medicine by enabling physicians to establish therapies designed for each patient, i.e., personalized medicine.

For example, many types of cancer that are distinct at the cellular level nevertheless have similar symptoms. Because symptoms may be similar in one genetic type of cancer and another, it is important to know everything possible about cancer genes and their interactions in prescribing an effective treatment.

As another example, physicians will be able to use a molecular/genomic test to help select the most effective drug with the minimum number of side effects. As a result,

this approach should benefit the patient with more customized care, reduced length of illness, and, ultimately, a better and longer life.

Besides healthcare, the field of agriculture is also likely to benefit from genomic research. The ability to diagnose plant and animal diseases and develop treatments targeted against those diseases should produce better agricultural products and improve yields. For example, the comparison of genetic information from disease-or pest-resistant plant strains with non-resistant strains and the use of selective breeding programs for favorable traits will significantly increase the number and success of new strains available to various agricultural areas around the world. This has major implications for not only increasing the quantity of food but also its nutritional quality.

Other fields that will likely derive important benefits from genomic information include forensics, veterinary medicine, textile production, waste control, and environmental remediation.

A significant impediment to achieving any of the foregoing expected beneficial results of genomic research is the sheer size of the amount of genomic information to be sifted through and studied. Without exaggeration, the amount of raw nucleic sequencing data available to be sifted through and studied is unimaginably vast. Moreover, the body of data gets bigger every day (literally), since newly sequenced strands of DNA are documented on an ongoing basis. Because of the vastness of the genomic data, it is stored, handled, and manipulated via computers.

Bioinformatics is the use of computers to retrieve, process, and analyze biological information. This field of data processing is now considered essential for drug discovery and development. Scientists are augmenting traditional "wet" biology with quantitative analyses, database comparisons, and computational algorithms. In this way, biology

research, at least preliminarily, is conducted in a virtual environment before the scientist sets foot in the laboratory. Bioinformatic tools and services assist pharmaceutical and biotechnology researchers with all phases of drug discovery and development including gene discovery, understanding disease pathways, identifying new disease targets and the discovery and correlation of gene sequence variation to disease.

Unfortunately, conventional means of access to genomic information does not provide for comprehensive and easy access so that the data can be analyzed or studied in a computationally transparent manner. Genomic information is accumulated in a wide variety of databases, some public (freely available), some commercial (available for a price), and some proprietary ("in-house" resources that are not shared with outsiders).

Referring to **Fig. 1**, genomic databases, both public **102** and private **104**, as well as search tools **106** are available online. A user **110** uses an interface device **112** to access the databases **102**, **104** and search tools **106**. The interface device **112** communicates with a data access portal site **120** via an Internet connection **130**. The portal **120** makes connections to the databases **102**, **104** and search tools **106** via the Internet **140**. Another operational mode for the user **110** to access the genomic research resources **102**, **104**, **106** is serially via an Internet connection without using the portal **120**. (This simplified mode of connection is not illustrated.) The interface device **112** is typically an implementation of a thin client browser application.

Each database is made up of data objects that contain biological data. The data objects differ from one database to the next in terms of the types of biological data they contain, and in terms of their formats. Thus, studying data from plural databases requires the researcher to learn how to interpret the data as presented in the unique data structures and data contents of each particular database. This is a significant inconvenience.

Thus, what is needed is a bioinformatics visualization tool that will automatically interpret data from diverse genomic databases (each containing genomic data objects of varying formats) so that it is presented to a user in a predictable, easily-recognizable format.

Additionally, because of the disparate storage formats used by the different databases, many of the data objects to be analyzed using automated analysis tools must be converted from their native format into a format that is recognizable to the automated analysis tool to be used. This is a further inconvenience.

Thus, what is needed is a bioinformatics software tool that will automatically translate data from diverse genomic databases (each containing genomic data objects of varying formats) so that it is presented to analysis facilities according to a uniform format.

Furthermore, when it is discovered that a pair of data objects has a significant relationship to one another, there is no conventional mechanism (other than manually scribbling a note to oneself) for establishing a linking relationship between them. Such a mechanism is most conspicuously absent in the case where the pair of data objects are found in two entirely different databases.

Thus, what is needed is a software facility that enables a user to establish linking relationships between data objects, even when those data objects are drawn from diverse databases and have diverse formats and types of content.

Moreover, the existing way for accessing and navigating Hyper Text Markup Language (HTML) genomic data files provided by Internet database hosts or by a centralized Internet portal has crucial limitations. For one thing, there is no interactivity possible for manipulation and/or analysis of the data. This interactivity is crucial in terms of research effectiveness.

Thus, what is needed is a software facility that enables a user to interactively access and navigate genomic data files over the Internet, whether those files be in the form of HTML (as is typically done) or in other formats the data may appear.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a bioinformatics visualization tool that will automatically interpret data from diverse genomic databases (each containing genomic data objects of varying formats) so that it is presented to a user in a predictable, easily-recognizable format.

It is another object of the present invention to provide a bioinformatics software process and system that will automatically translate data from diverse genomic databases (each containing genomic data objects of varying formats) so that it is presented to analysis facilities according to a uniform format.

It is yet another object of the present invention to provide a software process and system that enables a user to establish linking relationships between data objects, even when those data objects are drawn from diverse databases and have diverse formats and types of content.

It is still another object of the present invention to provide a software process and system that enables a user to interactively access and navigate HTML (or other format) genomic data files over the Internet.

It is a further object of the present invention to provide a business process for enabling seamless access and interactivity with plural genomic databases, even in the case where one or more of those databases is a commercial (i.e., fee for access) database.

Some of the above objects are made possible by a data processing system that is in electronic communication with a local genomic database system and with one or more

remote genomic database systems. The data processing system includes a graphical user interface that enables a user to view genomic data objects graphically. It also includes a genomic data object linker and a linkable data object resolver that resolves one or more genomic data objects, which are linkable with respect to a subject genomic data object, from among data objects found in the local genomic database system and the one or more remote genomic database systems. A resolved genomic data object that is resolved by the linkable data object resolver is linked to the subject genomic data object by the data object linker, so that the resolved genomic data object and the subject genomic data object are each provided to the graphical user interface.

Some of the above objects are also made possible by a system for performing genomic research that is in electronic communication with a local genomic database system and with one or more remote genomic database systems. The system includes a means for presenting a user with a graphical view of genomic data objects, as well as a means for linking genomic data objects to one another. The system further includes a means for resolving a genomic data object with respect to a subject genomic data object, from among genomic data objects found in the local genomic database system and the one or more remote genomic database systems. A resolved genomic data object that is resolved by the means for resolving is linked to the subject genomic data object by the means for linking, so that the resolved genomic data object and the subject genomic data object are each provided to the means for printing.

Another way that some of the above objects are made possible is by a method of performing genomic research with respect to a subject genomic data object, using a local genomic database and one or more remote genomic databases as resources. The method includes the act of resolving a linkable genomic data object, with respect to the subject

genomic data object, from among the local genomic database and the one or more remote genomic databases, regardless of the data formats of the genomic data objects. Additionally, the method includes the act of linking the subject genomic data object with the linkable genomic data object to form a set of linked genomic data objects. Furthermore, the method includes the act of storing the set of linked genomic data objects in the local genomic database.

Still another way that some of the above objects are made possible is by a computer system for use in genomic research that implements a method as described above.

One of the above objects is made possible by a novel technical information model and implementation facility. This is accomplished by combining genomic visualization and analysis tools with the ability to link genomic data objects across (and within) plural databases.

Another of the above objects is made possible by a method of administering access to a plurality of genomic databases, where the genomic databases include a local genomic database, a public genomic database, and a commercial genomic database. The method includes a step of resolving linkable data objects with respect to a subject data object, and a further step of linking the resolved data objects to the subject genomic data object. The linkable data objects that are resolved from public genomic databases are resolved regardless of the data formats of genomic data objects stored therein, and without restriction as to access costs. The linkable data objects that are resolved from commercial genomic databases are resolved regardless of data formats of genomic data objects stored therein, and are resolved subject to applicable, predetermined access agreements for the commercial genomic databases.

Still another way that some of the above objects are made possible is by a computer program product for enabling a computer to administer access to a plurality of genomic databases. The plurality of genomic databases includes a local genomic database, a public genomic database, and a commercial genomic database. The computer program product is software instructions for enabling the computer to perform predetermined operations, and a computer readable medium embodying the software instructions. The predetermined operations including the acts according to the methods discussed above.

One aspect of the present invention is a local database for storing genomic data, such as nucleic acid sequences, amino acid sequences, oligonucleotides, results of Basic Local Aligned Search Tool (BLAST) searches, and entries from medical databases such as MEDLINE.

Another aspect of the present invention is a visualization and analysis facility. Visualization and analysis is provide, preferably via dialog boxes, so that parameters may be set for BLAST searches; so that text-based searching may be made of a sequence database, and so that bibliographic search may be done of a database. A linker is included for linking together nucleic acid sequences, amino acid sequences, BLAST search results, MEDLINE entries, etc. All types of genomic/biomedical information are visualized via a graphical interface that incorporates viewer and editor components.

Still another aspect of the present invention is an Internet connector having a programming module that resolves links between database objects. This connector looks either in the local database or in a remote Internet server to obtain a data object being sought.

BRIEF DESCRIPTION OF THE DRAWING

Additional objects and advantages of the present invention will be apparent in the following detailed description read in conjunction with the accompanying drawing figures.

Fig. 1 illustrates a conventional configuration of using a browser to access Internet database resources via a portal.

Fig. 2 illustrates integrated access to Internet database resources according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

One way to view the present invention is that the traditional approach of using a thin client browser and a thin portal (refer to Fig. 1) to provide a researcher with access to genomic research resources is abandoned. In its place, applicants have discovered the increased effectiveness of a system that provides an interactive interface with the resources. This interactivity is crucial for increasing the effectiveness of genomic research. Another way to view the present invention is as a technical information model and implementation facility. The interactivity aspect of the invention is provided by the combination of linking of data objects and its visualization and analysis aspects.

In addition to a local database for storing intermediate and finalized results, the present invention has a visualization and analysis aspect. Visualization is provide in an advanced form that shows the sequences and other molecules in graphical presentations that are intuitively appealing to the human perception. In addition to interactivity, the ability for the user to easily integrate (i.e., link) data and store the integrated data in the local database is very helpful.

An additional functionality that is provided in a preferred embodiment of the present invention are full-scale analysis tools and algorithms directly available at the user

interface rather than remotely over the Internet. Full-scale analysis allows DNA to be evaluated in view of, protein, enzyme, and oligos data sets, BLAST results, MEDLINE Entrez data, and amino acids.

An example of a software product that can provide the visualization and analysis aspects of the present invention is the VectorNTI ™ product of InforMax, Inc. of North Bethesda, Maryland.

Another aspect according to a preferred embodiment of the present invention is the use of a resolving system that reaches out to plural databases over the Internet (e.g., NCBI, Entrez, PubMed, SRS) to provide integrated database searches.

Referring to **Fig. 2**, databases, both public **202** and private **204**, and various research tools **206** are available online. Also available as a research resource are the previous research results and other proprietary data that user **210** stores in a local database system **208**. As in the prior art, the Internet **220** is used as a communication medium to access the various remote resources **202, 204, 206**. However, in contrast with the prior art, an entirely different set of tools is used for conducting research.

The user **210** utilizes a user interface **230** that includes a visualization system **232** and a data set linking facility **234** (hereinafter "linker" for short). The linker **234** provides for integration of data sets that the user deems to be worthy of being associated with one another for further study in relation to one another. Data sets so linked may be more closely examined or analyzed using analysis tools and algorithms **240**. Examples of useful analysis tools and algorithms to include for use with the present invention are BioPlot™, AlignX™, and ContigExpress™, which are all products of InforMax, Inc. of North Bethesda, MD. A number of other available analysis tools, such as BLAST may also be usefully employed.

This examination and analysis produces results that may themselves be linked to the data sets from which they were derived. The local database **208** is used to store the integrated data sets and results for later study by the user **210**, or his or her colleagues. The later study may be in the form of additional computer analysis, or in a biology laboratory if the results are deemed to be sufficiently promising.

Candidates for linking are identified by the linkable data object resolving system **250**. The linkable data object resolving system **250** connects via the Internet **220** to access any of various search tools **206** and databases **202, 204** to search for data objects that are relevant to a subject data object that the user **210** has identified as being of interest. The resolving system **250** does not establish links. Rather, the resolving system **250** identifies data objects from the enormously vast collections of data that are available for inspection over the Internet that should have a reasonable probability of being relevant to the subject data object.

To limit and guide the searching by the resolving system **250**, the permissioning and accounting module **260** directs the resolving system **250** to access only databases that are public **202** or those commercial databases **204** for which access agreements have been established. The accounting aspect of the permissioning and accounting module **260** keeps records of access times, durations, and authorizations regarding the use of the commercial databasses **204**.

In addition to the apparatus aspects of the present invention, methods also form some aspects of the invention. A method of performing genomic research according to the present invention is accomplished, first, by resolving a linkable genomic data object, and then, by linking the linkable genomic data object with a subject genomic data object to form a set of linked genomic data objects. The resolving act is performed with respect to

the subject genomic data object, using one or more remote genomic databases as resources. Optionally, a local genomic database is also used in resolving a linkable genomic data object. The act of resolving is performed regardless of the data formats of the genomic data objects as they may be found in the various databases. After the genomic data objects are linked, they are preferably stored as a set of linked genomic data objects in the local genomic database.

Another aspect of the present invention is that it represents a business process wherein predetermined access agreements for commercial databases are used to guide research steps so that accessing of these databases is entirely seamless from the point of view of the researcher/user employing the process. This results in a process of administering access to a plurality of genomic databases, public, commercial, as well as local (possibly proprietary). The process includes a step of resolving linkable data objects with respect to a subject data object, and a further step of linking the resolved data objects to the subject genomic data object. The linkable data objects that are resolved from public genomic databases are resolved regardless of the data formats of genomic data objects stored therein, and without restriction as to access costs. The linkable data objects that are resolved from commercial genomic databases are resolved regardless of data formats of genomic data objects stored therein, and are resolved subject to applicable, predetermined access agreements for the commercial genomic databases.

The present invention has been described in terms of preferred embodiments, however, it will be appreciated that various modifications and improvements may be made to the described embodiments without departing from the scope of the invention. The scope of the invention is limited only by the appended claims.

WHAT IS CLAIMED IS:

1. A data processing system that is in communication with a local database system and that is in communication with one or more remote database systems, the data processing system comprising:

a graphical user interface enabling a user to view data objects graphically;

a data object linker;

a linkable data object resolver that resolves one or more data objects, which are linkable with respect to a subject data object, from among data objects found in the local database system and the one or more remote database systems;

wherein a resolved data object resolved by the linkable data object resolver is linked to the subject data object by the data object linker, so that the resolved data object and the subject data object are each provided to the graphical user interface.

2. A data processing system that is in electronic communication with a local genomic database system and that is in electronic communication with one or more remote genomic database systems, the data processing system comprising:

a graphical user interface enabling a user to view genomic data objects graphically;

a genomic data object linker;

a linkable data object resolver that resolves one or more genomic data objects, which are linkable with respect to a subject genomic data object, from among data objects found in the local genomic database system and the one or more remote genomic database systems;

wherein a resolved genomic data object resolved by the linkable data object resolver is linked to the subject genomic data object by the data object linker, so that the

resolved genomic data object and the subject genomic data object are each provided to the graphical user interface.

3. The data processing system of claim 2, wherein the data processing system communicates with the one or more remote genomic database systems via a network.

4. The data processing system of claim 2, wherein the data processing system communicates with the one or more remote genomic database systems via a network of networks.

5. The data processing system of claim 2, wherein the data processing system communicates with the one or more remote genomic database systems via the Internet.

6. The data processing system of claim 2, wherein the resolved genomic data object and the subject genomic data object are each of a data type selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of a BLAST search, and medical data.

7. The data processing system of claim 6, wherein the resolved genomic data object and the subject genomic data object are of different data types.

8. The data processing system of claim 2, wherein the local genomic database system and the one or more remote genomic database systems each contain data objects of types that are selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, and medical data.

9. The data processing system of claim 2, wherein the graphical user interface has the capability to graphically depict nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, and medical data.

10. The data processing system of claim 2, wherein the genomic data object linker links genomic data objects that are of differing data types.

11. A system for performing genomic research, the system being in electronic communication with a local genomic database system and that is in electronic communication with one or more remote genomic database systems, the system comprising:

means for presenting a user with a graphical view of genomic data objects;

means for linking genomic data objects to one another; and

means for resolving a genomic data object with respect to a subject genomic data object, from among genomic data objects found in the local genomic database system and the one or more remote genomic database systems;

wherein a resolved genomic data object resolved by the means for resolving is linked to the subject genomic data object by the means for linking, so that the resolved genomic data object and the subject genomic data object are each provided to the means for printing.

12. The system for performing genomic research of claim 11, wherein the system communicates with the one or more remote genomic database systems via the Internet.

13. The system for performing genomic research of claim 11, wherein the resolved genomic data object and the subject resolved genomic data object are each of a data type

selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of a BLAST search, and medical data.

14. The system for performing genomic research of claim 13, wherein the resolved genomic data object and the subject resolved genomic data object are of different data types.

15. The system for performing genomic research of claim 11, wherein the local genomic database system and the one or more remote genomic database systems each contain data objects of types that are selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, and medical data.

16. The system for performing genomic research of claim 11, wherein the means for presenting has the capability to graphically depict nucleic acid sequences, amino acid sequences, olgionuclcotides, results of BLAST searches, and medical data.

17. The system for performing genomic research of claim 11, wherein the means for linking links genomic data objects that are of differing data types.

18. A computer system adapted to genomic research with respect to a subject genomic data object, using a local genomic database and remote genomic databases as resources, the computer system comprising:

a processor, and

a memory, in electronic communication with the processor, including software instructions adapted to enable the computer system to perform the steps of:

resolve a linkable genomic data object, with respect to the subject genomic data object, from among the local genomic database and the remote genomic databases, regardless of the data formats of the genomic data objects;

link the subject genomic data object with the linkable genomic data object to form a set of linked genomic data objects; and

store the set of linked genomic data objects in the local genomic database.

19. A method of performing genomic research with respect to a subject genomic data object, using a local genomic database and one or more remote genomic databases as resources, the method comprising:

resolving a linkable genomic data object, with respect to the subject genomic data object, from among the local genomic database and the one or more remote genomic databases, regardless of the data formats of the genomic data objects;

linking the subject genomic data object with the linkable genomic data object to form a set of linked genomic data objects; and

storing the set of linked genomic data objects in the local genomic database.

20. The method of performing genomic research of claim 19, wherein the resolved genomic data object and the subject genomic data object are each of a data type selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of a BLAST search, and medical data.

21. The method of performing genomic research of claim 20, wherein the resolved genomic data object and the subject genomic data object are of different data types.

22. The method of performing genomic research of claim 19, wherein the local genomic database system and the one or more remote genomic database systems each contain data objects of types that are selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, and medical data.

23. The method of performing genomic research of claim 19, the method further comprising:

providing a graphical user interface to graphically depict the subject genomic data object and the resolved genomic data object, regardless of whether they are nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, or medical data.

24. The method of performing genomic research of claim 19, wherein the act of linking links genomic data objects that are of differing data types.

25. A method of administering access to a plurality of genomic databases, the plurality of genomic databases including a local genomic database, a public genomic database, and a commercial genomic database, the method comprising:

resolving one or more linkable data objects with respect to a subject data object; and

linking the one or more resolved data objects to said subject genomic data object;

wherein the one or more linkable data objects are resolved from public genomic databases, regardless of data formats of genomic data objects stored therein, and without restriction as to access costs; and

wherein the one or more linkable data objects are resolved from commercial genomic databases, regardless of data formats of genomic data objects stored therein, and subject to predetermined access agreements for the commercial genomic databases.

26. The method of administering access to a plurality of genomic databases recited in claim 25, wherein the resolved genomic data object and the subject genomic data object are each of a data type selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of a BLAST search, and medical data.

27. The method of administering access to a plurality of genomic databases recited in claim 26, wherein the resolved genomic data object and the subject genomic data object are of different data types.

28. The method of administering access to a plurality of genomic databases recited in claim 25, wherein the local genomic database system and the one or more remote genomic database systems each contain data objects of types that are selected from the group consisting of: nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, and medical data.

29. The method of administering access to a plurality of genomic databases recited in claim 25, the method further comprising:

providing a graphical user interface to graphically depict the subject genomic data object and the resolved genomic data object, regardless of whether they are nucleic acid sequences, amino acid sequences, olgionucleotides, results of BLAST searches, or medical data.

30. The method of administering access to a plurality of genomic databases recited in claim 25, wherein the act of linking links genomic data objects that are of differing data types.

31. A computer program product for enabling a computer to perform genomic research with respect to a subject genomic data object, using a local genomic database and remote genomic databases as resources, the computer program product comprising:

software instructions for enabling the computer to perform predetermined operations, and

a computer readable medium embodying the software instructions;

the predetermined operations including the acts of:

resolve a linkable genomic data object, with respect to the subject genomic data object, from among the local genomic database and the remote genomic databases, regardless of the data formats of the genomic data objects;

link the subject genomic data object with the linkable genomic data object to form a set of linked genomic data objects; and

store the set of linked genomic data objects in the local genomic database.

32. A computer program product for enabling a computer to administer access to a plurality of genomic databases, the plurality of genomic databases including a local genomic database, a public genomic database, and a commercial genomic database, the computer program product comprising:

software instructions for enabling the computer to perform predetermined operations, and

a computer readable medium embodying the software instructions;

the predetermined operations including the acts of:

resolve one or more linkable data objects with respect to a subject data object; and

linking the one or more resolved data objects to said subject genomic data object;

wherein the one or more linkable data objects are resolved from public genomic databases, regardless of data formats of genomic data objects stored therein, and without restriction as to access costs; and

wherein the one or more linkable data objects are resolved from commercial genomic databases, regardless of data formats of genomic data objects stored therein, and subject to predetermined access agreements for the commercial genomic databases.

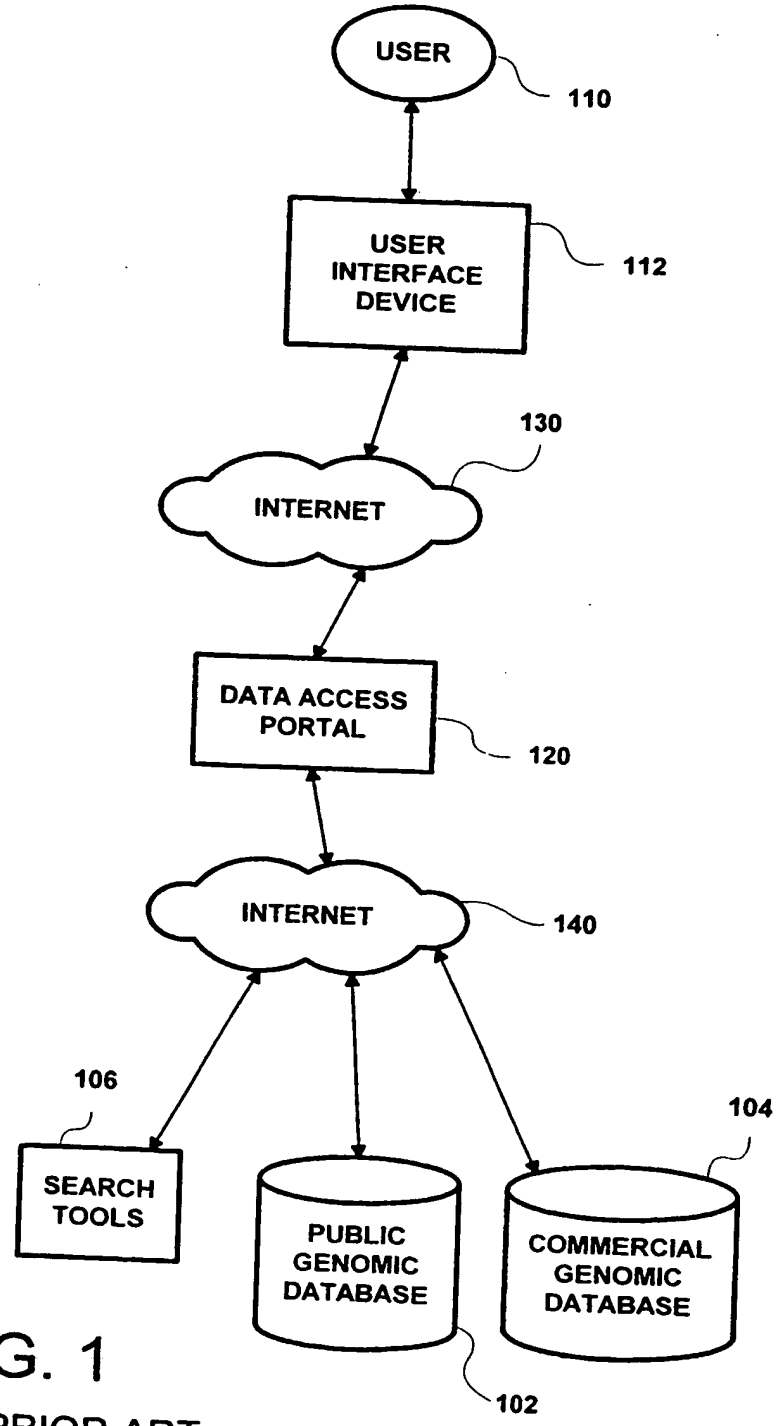


FIG. 1
PRIOR ART

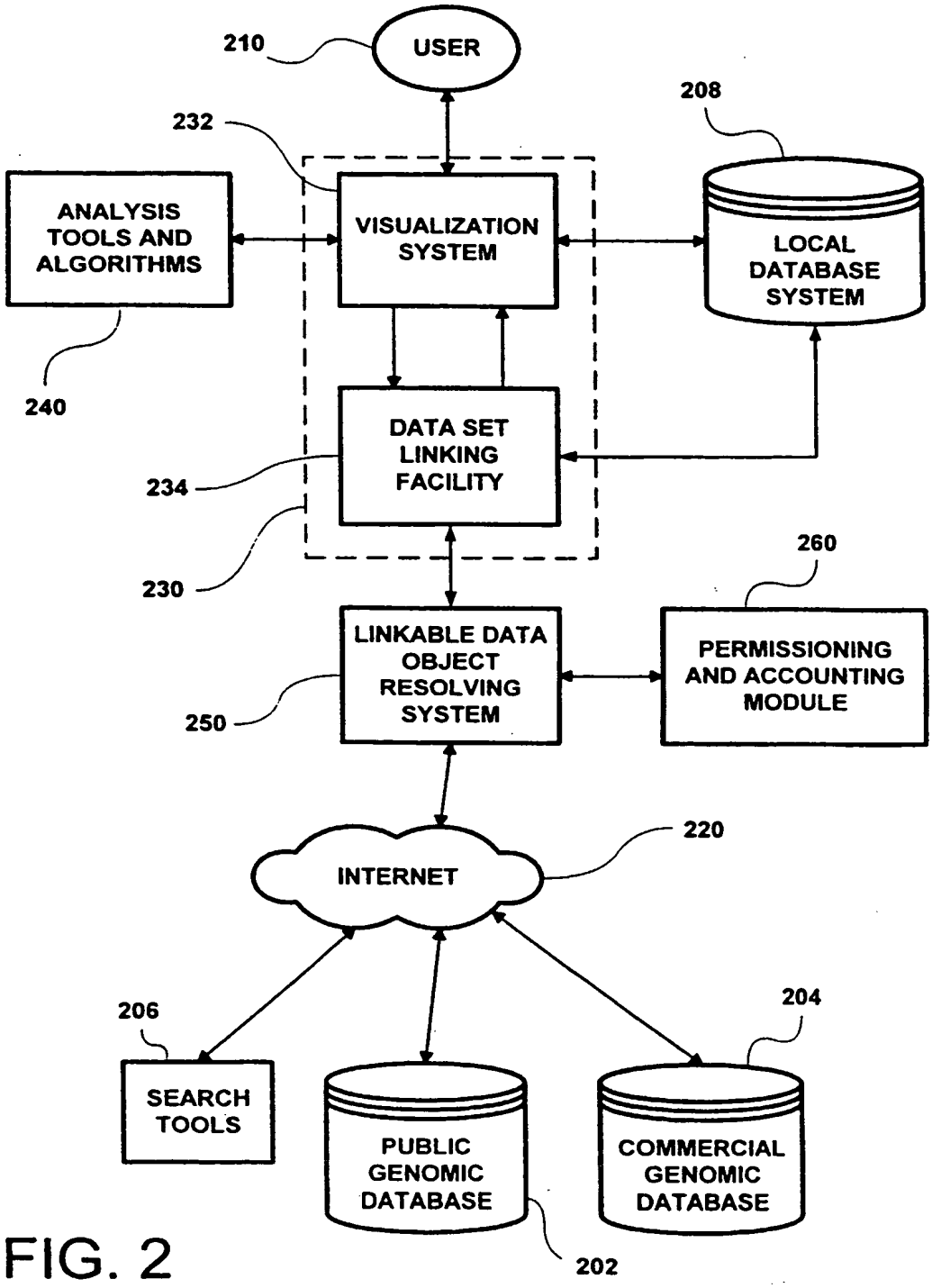

210
USER
208
232
ANALYSIS TOOLS AND ALGORITHMS
VISUALIZATION SYSTEM
LOCAL DATABASE SYSTEM
240
DATA SET LINKING FACILITY
234
230
260
LINKABLE DATA OBJECT RESOLVING SYSTEM
PERMISSIONING AND ACCOUNTING MODULE
250
220
INTERNET
206
204
SEARCH TOOLS
PUBLIC GENOMIC DATABASE
COMMERCIAL GENOMIC DATABASE
202

FIG. 2

INTERNATIONAL SEARCH REPORT

Internati pplication No
PCT/US 01/02527

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, EPO-Internal, PAJ, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DAVIDSON S.B. ET AL.: "BioKleisli: a digital library for biomedical researchers" INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES, vol. 1, no. 1, April 1997 (1997-04), pages 36-53, XP002167355 Germany paragraph '04.4! - paragraph '0005!; figure 6 abstract --- -/--	1-32

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 May 2001	29/05/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Deane, E

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internati Application No
PCT/US 01/02527

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	ACHARD F. : "Automatic generation of links between heterogeneous genomic databases" PROC. FIRST INT. SYMPOSIUM ON INTELLIGENCE IN NEURAL AND BIOLOGICAL SYSTEMS, 29 - 31 May 1995, pages 78-83, XP002167356 Herndon VA, USA abstract page 81, left-hand column, line 1 -page 82, right-hand column, line 12 ---	1,2,11, 18,19, 25,31,32
A	YEE D.P. ET AL.: "Automated clustering and assembly of large EST collections" PROC. 6TH. INT. CONF. ON INTELLIGENT SYSTEMS FOR MOELCULAR BIOLOGY, 28 June 1998 (1998-06-28) - 1 July 1998 (1998-07-01), pages 203-211, XP000994591 Montreal, Canada abstract ---	1,2,11, 18,19, 25,31,32
A	MOUSHENG XU ET AL.: "Associated Biological Information Retrieval From Distributed Databases" PROC. 1998 ACM CIKM 7TH. INT. CONF. ON INFORMATION AND KNOWLEDGE MANAGEMENT, 3 - 7 November 1998, pages 193-200, XP002167357 Bethesda, MD, USA page 193, left-hand column, line 1 -page 195, left-hand column, line 31 -----	1,2,11, 18,19, 25,31,32